Aalto University
School of Electrical Engineering

ELEC-E7450
Performance Analysis

# Basic probability theory

Samuli Aalto
Department of Communications and Networking

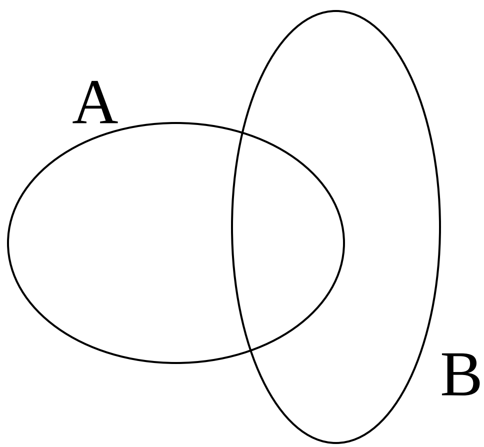
# Contents

# Sample space, sample points, events

- Sample space $\Omega$ is the set of all possible sample points $\omega \in \Omega$
- Events $A,B,C,\ldots \subset \Omega$ are measurable subsets of the sample space $\Omega$
- Let $\mathcal{F}$ denote the set of all events $A$, which constitutes a $\sigma$-algebra
  - Sure event: The whole sample space $\Omega \in \mathcal{F}$
  - Impossible event: The empty set $\emptyset \in \mathcal{F}$
  - Union "A or B": $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ or } \omega \in B\} \in \mathcal{F}$
  - Intersection "A and B": $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ and } \omega \in B\} \in \mathcal{F}$
  - Complement "not A": $A^c = \{\omega \in \Omega \mid \omega \notin A\} \in \mathcal{F}$
  - Events $A$ and $B$ are disjoint if $A \cap B = \emptyset$
  - A set of events $\{B_1, B_2, \ldots\}$ is a partition of event $A$ if
    - $(i)$ $B_i \cap B_j = \emptyset$ for all $i \neq j$
    - $(ii)$ $\cup_i B_i = A$

# Probability

- Probability of event $A$ is denoted by $P(A) \in [0,1]$
  - Probability measure $P$ is thus
    a real-valued set function defined on the set $\mathcal{F}$ of events, $P\colon \mathcal{F} \rightarrow [0,1]$
- **Properties:**
  - $(i)$ $0 \leq P(A) \leq 1$
  - $(ii)$ $P(\varnothing) = 0$
  - $(iii)$ $P(\Omega) = 1$
  - $(iv)$ $P(A^c) = 1 - P(A)$
  - $(v)$ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $(vi)$ $A \cap B = \varnothing \Rightarrow \quad P(A \cup B) = P(A) + P(B)$
  - $(vii)$ $\{B_i\}$ is a partition of $A \Rightarrow \ P(A) = \Sigma_i P(B_i)$
  - (viii) $A \subset B \Rightarrow P(A) \leq P(B)$



# Conditional probability

- Assume that $P(B) > 0$
- **Definition:**
  The conditional probability of event $A$ given that event $B$ occurred is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- It follows that

$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$$

# Theorem of total probability

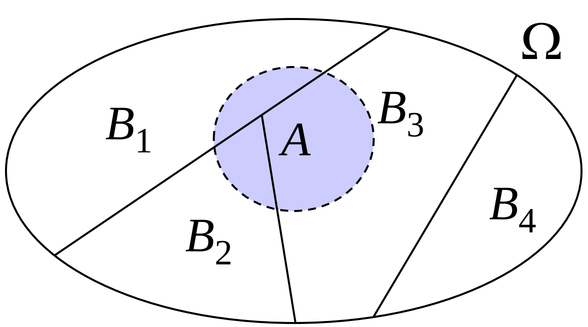
- Let $\{B_i\}$ be a partition of the sample space $\Omega$
- It follows that $\{A \cap B_i\}$ is a partition of event $A$. Thus (by slide 4)

$$P(A) \overset{(vii)}{=} \sum_i P(A \cap B_i)$$

- Assume further that $P(B_i) > 0$ for all $i$. Then (by slide 5)

$$P(A) = \sum_i P(B_i) P(A \mid B_i)$$

- This is the theorem of total probability



# Statistical independence of events

- **Definition:**
  Events $A$ and $B$ are independent if

$$P(A \cap B) = P(A)P(B)$$

- If $A$ and $B$ are independent, then

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

- Correspondingly:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$$

# Random variables

- **Definition:**
  Real-valued random variable $X$ is a measurable function defined on the sample space $\Omega$, $X: \Omega \to \Re$
  – Each sample point $\omega \in \Omega$ is associated with a real number $X(\omega)$
- Measurability means that all sets of type

$$\{X \leq x\} := \{\omega \in \Omega \mid X(\omega) \leq x\} \subset \Omega$$

  belong to the set $\mathcal{F}$ of events, i.e.,

$$\{X \leq x\} \in \mathcal{F}$$

- The probability of such an event is denoted by $P\{X \leq x\}$
- **Notation:**
  Capital Letters (such as $X$) refer to random variables, while small letters (such as $x$) refer to their values

# Indicators of events

- Let $A \in \mathcal{F}$ be an arbitrary event
- **Definition:**
  The indicator of event $A$ is a random variable defined by

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

- Clearly:

$$P\{1_A = 1\} = P(A)$$

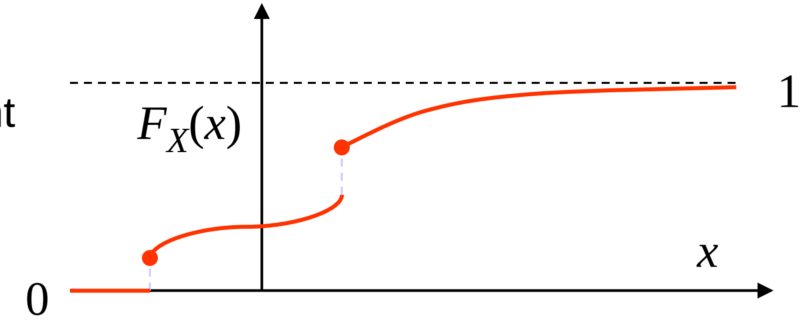
$$P\{1_A = 0\} = P(A^c) = 1 - P(A)$$

# Cumulative distribution function

- **Definition:**
  The cumulative distribution function (CDF) of a random variable $X$ is a function $F_X\colon \Re \to [0,1]$ defined as follows:

$$F_X(x) := P\{X \leq x\}$$

- CDF determines the distribution of the random variable, i.e.,
  - the probabilities $P\{X \in B\}$, where $B \subset \Re$ and $\{X \in B\} \in \mathcal{F}$
- **Properties:**
  - $(i)$ $F_X$ is non-decreasing
  - $(ii)$ $F_X$ is continuous from the right
  - $(iii)$ $F_X(-\infty) = 0$
  - $(iv)$ $F_X(\infty) = 1$



# Statistical independence of random variables

- **Definition:**
  Random variables $X$ and $Y$ are independent if for all $x$ and $y$

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\} P\{Y \leq y\}$$

- **Definition:**
  Random variables $X_1, \ldots, X_n$ are (totally) independent if for all $i$ and $x_i$

$$P\{X_1 \leq x_1, \ldots, X_n \leq x_n\} = P\{X_1 \leq x_1\} \cdots P\{X_n \leq x_n\}$$

- **Definition:**
  Random variables $X_1, \ldots, X_n$ are IID if they are independent and identically distributed
- **Note:**
  If $X$ and $Y$ are independent, then also random variables $f(X)$ and $g(Y)$ are independent for any (measurable) functions $f(x)$ and $g(y)$

# Contents

- Basic concepts
- Discrete random variables
- Conditional expectation and variance
- Discrete distributions (count distributions)
- Continuous random variables
- Continuous distributions (time distributions)
- Other distributions and random variables

# Discrete random variables

- **Definition:**
  Set $A \subset \Re$ is called discrete if it is
  - finite, $A = \{x_1, \ldots, x_n\}$, or
  - countably infinite, $A = \{x_1, x_2, \ldots\}$
- **Definition:**
  Random variable $X$ is discrete if
  there is a discrete set $S_X \subset \Re$ such that

$$P\{X \in S_X\} = 1$$

- It follows that
  - $P\{X = x\} \geq 0$ for all $x \in S_X$
  - $P\{X = x\} = 0$ for all $x \notin S_X$
- **Definition:**
  The set $S_X$ is called the value space of $X$

# Point probabilities

- Let $X$ be a discrete random variable
- The distribution of $X$ is determined by the point probabilities $p_i$,

$$p_i := P\{X = x_i\}, \qquad x_i \in S_X$$

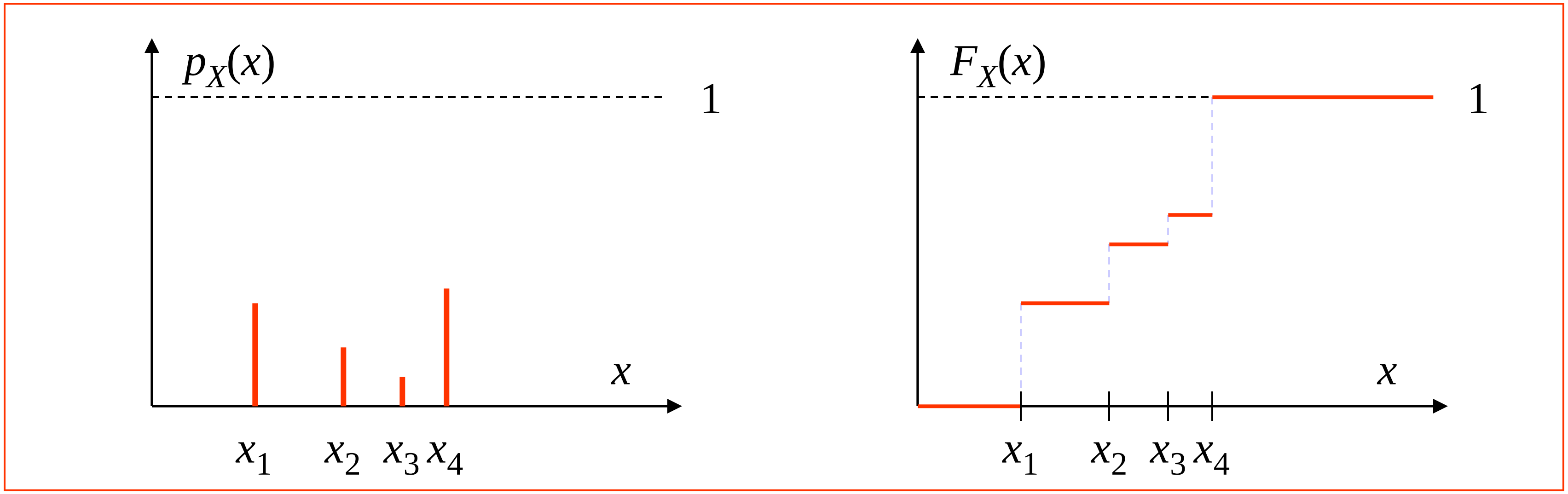
- **Definition:**
  The probability mass function (PMF) of $X$ is defined by

$$p_X(x) := P\{X = x\} = \begin{cases} p_i, & x = x_i \in S_X \\ 0, & x \notin S_X \end{cases}$$

- CDF is in this case a step function:

$$F_X(x) = P\{X \le x\} = \sum_{i: x_i \le x} p_i$$

# Example




probability mass function (PMF)  cumulative distribution function (CDF)

$$S_X = \{x_1, x_2, x_3, x_4\}$$

# Expectation

- **Definition:**
  The expectation (mean value) of a discrete random variable $X$ is defined by

$$E[X] := \sum_{x \in S_X} P\{X = x\} \cdot x$$

  – **Note**: Expectation of an indicator: $E[1_A] = P\{1_A = 1\} = P(A)$

- **Properties:**
  – $(i)$ $c \in \Re \Rightarrow E[cX] = cE[X]$
  – $(ii)$ $E[X+Y] = E[X] + E[Y]$
  – $(iii)$ $X$ and $Y$ independent $\Rightarrow E[XY] = E[X]E[Y]$

# Monotone Convergence Theorem

- **Theorem:**
  If $X_i \geq 0$ for all $i$, then

$$E[\sum_{i=1}^{\infty} X_i] = \sum_{i=1}^{\infty} E[X_i]$$

# Variance

- **Definition:**
  The variance of $X$ is defined by

$$D^2[X] := \mathrm{Var}[X] := E[(X - E[X])^2]$$

- Useful formula:

$$D^2[X] = E[X^2] - E[X]^2$$

- **Properties:**
  - $(i)$ $c \in \Re \Rightarrow D^2[cX] = c^2 D^2[X]$
  - $(ii)$ $X$ and $Y$ independent $\Rightarrow D^2[X + Y] = D^2[X] + D^2[Y]$

# Other distribution related parameters

- **Definition:**
  The standard deviation of $X$ is defined by

$$D[X] := \sqrt{D^2[X]}$$

- **Definition:**
  The coefficient of variation of $X \geq 0$ is defined by

$$C[X] := \frac{D[X]}{E[X]}$$

- **Definition:**
  The *k*th moment, $k = 1, 2, \ldots$, of $X$ is defined as

$$E[X^k] = \sum_x P\{X = x\} \cdot x^k$$

# Average of IID random variables

- Let $X_1, \ldots, X_n$ be independent and identically distributed (IID) with mean $\mu$ and variance $\sigma^2$
- Denote the average (sample mean) as follows:

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Then

$$E[\bar{X}_n] = \mu$$

$$D^2[\bar{X}_n] = \frac{\sigma^2}{n}$$

$$D[\bar{X}_n] = \frac{\sigma}{\sqrt{n}}$$

# Law of large numbers (LLN)

- Let $X_1, \ldots, X_n$ be independent and identically distributed (IID) with mean $\mu$ and variance $\sigma^2$
- Weak law of large numbers: for all $\varepsilon > 0$

$$P\{|\bar{X}_n - \mu| > \varepsilon\} \to 0$$

- Strong law of large numbers: with probability $1$

$$\bar{X}_n \to \mu$$

- It follows that for large values of $n$

$$\bar{X}_n \approx \mu$$

# Theorem of total probability

- Let $X$ be a random variable. If $Y$ is a discrete random variable, then

$$P\{X \le x\} = \sum_j P\{Y = y_j\} P\{X \le x \mid Y = y_j\}$$

- Application of the theorem of total probability

# Contents

- Basic concepts
- Discrete random variables
- Conditional expectation and variance
- Discrete distributions (count distributions)
- Continuous random variables
- Continuous distributions (time distributions)
- Other distributions and random variables

# Conditional expectation

- **Definition:**
  Let $X$ and $Y$ be discrete random variables. The conditional expectation $E[X|Y]$ of $X$ (conditioned on $Y$) is a random variable defined by

$$E[X\,|\,Y] := f(Y)$$
$$f(y) := E[X\,|\,Y = y] := \sum_{x \in S_X} P\{X = x\,|\,Y = y\} \cdot x$$

- **Properties:**
  - $(i)$ $E[g(Y)\,X|Y] = g(Y)\,E[X|Y]$
  - $(ii)$ $E[X + Y|Z] = E[X|Z] + E[Y|Z]$
  - $(iii)$ $X$ and $Y$ independent $\Rightarrow E[X|Y] = E[X]$
  - (iv) $E[E[X|Y]] = E[X]$ (conditioning rule)

# Wald's equation

- Let $X_1, X_2, \dots$ be IID random variables with mean $E[X]$. In addition, let $N$ be another independent random variable taking values in $\{0,1,2,\dots\}$. The mean of the random sum $X_1 + \dots + X_N$ is given by Wald's equation

$$E[\sum_{i=1}^{N} X_i] = E[N]E[X]$$

- **Proof:**

$$\begin{aligned} E[\sum_{i=1}^{N} X_i] &= E[E[\sum_{i=1}^{N} X_i \mid N]] \\ &= E[N \cdot E[X]] \\ &= E[N] \cdot E[X] \end{aligned}$$

# Conditional variance

- **Definition:**
  Let $X$ and $Y$ be discrete random variables. The conditional variance of $X$, conditioned on $Y$, is a random variable defined by

$$D^2[X \mid Y] := E[(X - E[X \mid Y])^2 \mid Y]$$

- Useful formula:

$$D^2[X] = E[D^2[X \mid Y]] + D^2[E[X \mid Y]]$$

# Contents

# Bernoulli distribution

$$X \sim \text{Bernoulli}(p), \quad p \in (0,1)$$

- describes a simple random experiment (called Bernoulli trial) with two possible outcomes: success (1) and failure (0); cf. coin tossing
- success with probability $p$ (and failure with probability $1 - p$)

- Value space: $S_X = \{0,1\}$
- Point probabilities:

$$P\{X = 0\} = 1 - p, \quad P\{X = 1\} = p$$

- Mean value: $E[X] = (1 - p)\cdot 0 + p\cdot 1 = p$
- Second moment: $E[X^2] = (1 - p)\cdot 0^2 + p\cdot 1^2 = p$
- Variance: $D^2[X] = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)$

# Binomial distribution

$$X \sim \mathrm{Bin}(n, p), \quad n \in \{1,2,\ldots\}, p \in (0,1)$$

- number of successes in a finite sequence of IID Bernoulli trials; $X = X_1 + \ldots + X_n$ with $X_i \sim \mathrm{Bernoulli}(p)$
- $n$ = total number of experiments
- $p$ = probability of success in any single experiment

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

$$n! = n \cdot (n-1) \cdots 2 \cdot 1$$

- Value space: $S_X = \{0,1,\ldots,n\}$
- Point probabilities:

$$P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

- Mean value: $E[X] = E[X_1] + \ldots + E[X_n] = np$
- Variance: $D^2[X] = D^2[X_1] + \ldots + D^2[X_n] = np(1-p)$

# Geometric distribution

$$X \sim \text{Geom}(p), \quad p \in (0,1)$$

- number of successes until the first failure in a sequence of IID Bernoulli trials
- $p$ = probability of success in any single experiment

- Value space: $S_X = \{0,1,\ldots\}$
- Point probabilities:

$$P\{X = i\} = p^i(1-p)$$

- Mean value: $E[X] = \sum_i ip^i(1-p) = p/(1-p)$
- Second moment: $E[X^2] = \sum_i i^2p^i(1-p) = 2(p/(1-p))^2 + p/(1-p)$
- Variance: $D^2[X] = E[X^2] - E[X]^2 = p/(1-p)^2$

# Memoryless property

- Geometric distribution has so called memoryless property:
  for all $i,j \in \{0,1,...\}$

$$P\{X \geq i + j \mid X \geq i\} = P\{X \geq j\}$$

- **Proof:**

$$P\{X \geq i + j \mid X \geq i\} = \frac{P\{X \geq i + j\}}{P\{X \geq i\}} = \frac{p^{i+j}}{p^{i}} = p^{j} = P\{X \geq j\}$$

# Poisson distribution

$$X \sim \text{Poisson}(a), \quad a > 0$$

- the limit of binomial distribution as $n \to \infty$ and $p \to 0$ so that $np \to a$

- Value space: $S_X = \{0,1,\ldots\}$
- Point probabilities:

$$P\{X = i\} = \frac{a^i}{i!} e^{-a}$$

- Mean value: $E[X] = a$
- Second moment: $E[X(X-1)] = a^2 \Rightarrow E[X^2] = a^2 + a$
- Variance: $D^2[X] = E[X^2] - E[X]^2 = a$

# Properties

- ($i$) Sum: Let $X_1 \sim \text{Poisson}(a_1)$ and $X_2 \sim \text{Poisson}(a_2)$ be independent. Then

$$X_1 + X_2 \sim \text{Poisson}(a_1 + a_2)$$

- ($ii$) Random sample: Let $X \sim \text{Poisson}(a)$ denote the number of elements in a set, and $Y$ denote the size of a random sample of this set (each element taken independently with probability $p$). Then

$$Y \sim \text{Poisson}(pa)$$

- ($iii$) Random sorting: Let $X$ and $Y$ be as in ($ii$), and $Z = X - Y$. Then $Y$ and $Z$ are independent (given that $X$ is unknown) and

$$Z \sim \text{Poisson}((1-p)a)$$

# Contents

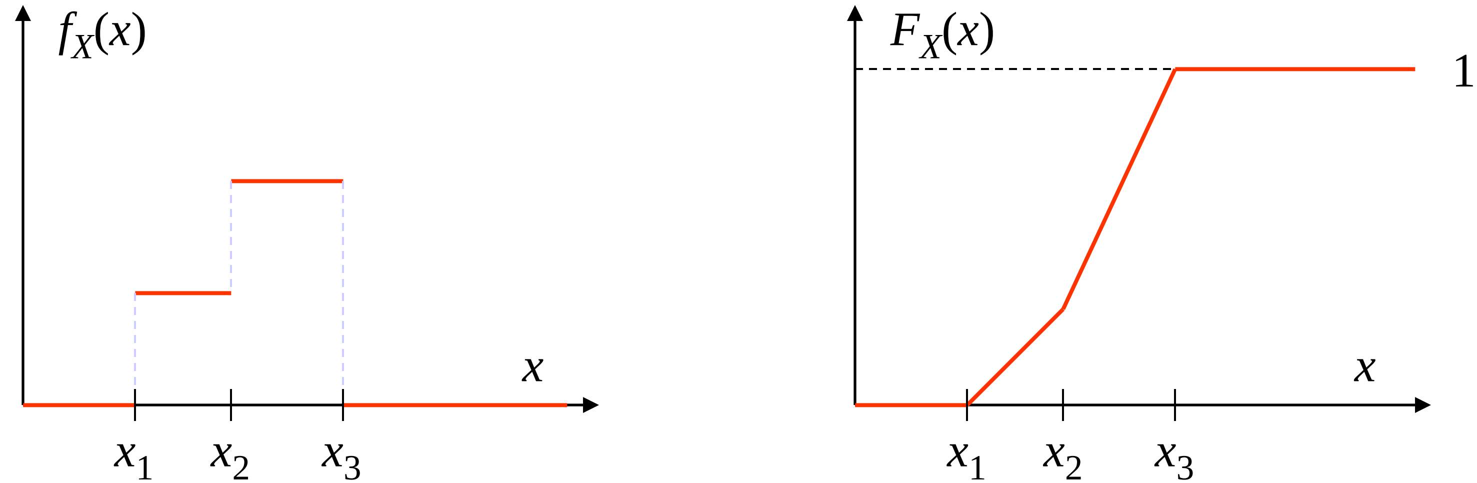
# Continuous random variables

- **Definition:**
  Random variable $X$ is continuous if there is an integrable function $f_X: \Re \to [0, \infty)$ such that for all $x \in \Re$

$$F_X(x) := P\{X \leq x\} = \int_{-\infty}^{x} f_X(y)\, dy$$

  - Function $f_X$ is called the probability density function (PDF)
  - Set $S_X$, where $f_X > 0$, is called the value space
- **Properties:**
  - $(i)$ $P\{X = x\} = 0$ for all $x \in \Re$
  - $(ii)$ $P\{a < X < b\} = P\{a \leq X \leq b\} = \int_a^b f_X(x)\, dx$
  - $(iii)$ $P\{X \in A\} = \int_A f_X(x)\, dx$
  - $(iv)$ $P\{X \in \Re\} = \int_{-\infty}^{\infty} f_X(x)\, dx = \int_{S_X} f_X(x)\, dx = 1$

# Example




probability density function (PDF) cumulative distribution function (CDF)

$$S_X = [x_1, x_3]$$

# Expectation and other distribution related parameters

- **Definition**:
  The expectation (mean value) of $X$ is defined by

$$E[X] := \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

- The expectation has the same properties as in the discrete case!
- The other distribution parameters (variance, standard deviation,...) are defined just as in the discrete case
  - These parameters have the same properties as in the discrete case

# Alternative mean value formula

- If $X \geq 0$ and $c \geq 0$, then

$$E[X] = \int_0^{\infty} P\{X > x\} dx$$

$$E[\min\{X, c\}] = \int_0^{c} P\{X > x\} dx$$

# Theorem of total probability

- Let $X$ be a random variable. If $Y$ is a continuous random variable, then

$$P\{X \leq x\} = \int_{-\infty}^{\infty} f_Y(y) P\{X \leq x \mid Y = y\}\, dy$$

- Application of the theorem of total probability

# Contents

- Basic concepts
- Discrete random variables
- Conditional expectation and variance
- Discrete distributions (count distributions)
- Continuous random variables
- Continuous distributions (time distributions)
- Other distributions and random variables

# From geometric to exponential distribution

- Assume that $X_n \sim \text{Geom}(1 - \mu/n)$ for some $\mu > 0$. Now

$$P\{X_n \geq nx\} = (1 - \frac{\mu}{n})^{nx} \to e^{-\mu x}$$

- Thus, the asymptotic CDF of the rescaled random variable $X_n/n$ is

$$F(x) = 1 - e^{-\mu x}$$

# Exponential distribution

$$X \sim \mathrm{Exp}(\mu), \quad \mu > 0$$

- continuous counterpart of the geometric distribution ("failure" prob. $\approx \mu dt$)
- $\mu$ = intensity (of an exponential phase)
- $P\{X \in (t,t+h] \mid X > t\} = \mu h + o(h)$, where $o(h)/h \to 0$ as $h \to 0$

- Value space: $S_X = (0,\infty)$
- PDF and CDF:

$$f_X(x) = \mu e^{-\mu x}, \quad x > 0$$

$$F_X(x) := P\{X \le x\} = 1 - e^{-\mu x}$$

# Moments

$$X \sim \text{Exp}(\mu), \quad \mu > 0$$

- Mean value: $E[X] = \int_0^\infty \mu x\, e^{-\mu x}\, dx = 1/\mu$
- Second moment: $E[X^2] = \int_0^\infty \mu x^2\, e^{-\mu x}\, dx = 2/\mu^2$
- Variance: $D^2[X] = E[X^2] - E[X]^2 = 1/\mu^2$
- Standard deviation: $D[X] = \sqrt{D^2[X]} = 1/\mu$
- Coefficient of variation: $C[X] = D[X]/E[X] = 1$

# Memoryless property and the residual lifetime

- Exponential distribution has so called **memoryless property**:
  for all $x,y \in (0,\infty)$

$$P\{X > x + y \mid X > x\} = P\{X > y\}$$

- In fact, only the exponential distribution has this property (among the continuous distributions)
- Consider a random interval of length $X \sim \mathrm{Exp}(\mu)$.
  Assume that we know that the interval is longer than $x$.
  Due to the memoryless property, the residual lifetime is also exponentially distributed with mean $1/\mu$:

$$\mathrm{MRL}(x) := E[X - x \mid X > x] = \frac{1}{\mu}$$

- Thus, the mean residual lifetime function $\mathrm{MRL}(x)$ is constant

# Hazard rate

- Consider a random interval of length $X \sim \mathrm{Exp}(\mu)$.
  Assume that we know that the interval is longer than $x$.
  What is the probability that it will end in an infinitesimal interval of length $h$ after time $x$?

$$P\{X \le x+h \mid X > x\} = P\{X \le h\} = 1 - e^{-\mu h}$$
$$= 1 - (1 - \mu h + \tfrac{1}{2}(\mu h)^2 - \ldots) = \mu h + o(h)$$

- Thus, in the limit ($h \to 0$), the ending probability per time unit (hazard rate) is constant:

$$h(x) := \lim_{h \to 0} \frac{1}{h} P\{X \le x+h \mid X > x\} = \mu$$

- Again, only the exponential distribution has this property

# Minimum of exponential random variables

- Let $X_1,\ldots, X_n$ be independent random variables with $X_i \sim \mathrm{Exp}(\mu_i)$. Then

$$X^{\min} := \min\{X_1,\ldots, X_n\} \sim \mathrm{Exp}(\mu_1 + \ldots + \mu_n)$$

since

$$P\{X^{\min} > x\} = P\{X_1 > x\}\ldots P\{X_n > x\} = e^{-(\mu_1+\ldots+\mu_n)x}$$

- In addition, we have

$$E[X^{\min}] = \frac{1}{\mu_1+\ldots+\mu_n}, \qquad P\{X^{\min} = X_i\} = \frac{\mu_i}{\mu_1+\ldots+\mu_n}$$

# Erlang distribution

$$X \sim \mathrm{Erl}(n, \mu), \quad \mu > 0$$

- IID exponential phases in a series; $X = X_1 + \ldots + X_n$, where $X_i \sim \mathrm{Exp}(\mu)$
- $n$ = total number of phases
- $\mu$ = intensity of any single phase

- Value space: $S_X = (0,\infty)$
- PDF and CDF:

$$f_X(x) = \mu \frac{(\mu x)^{n-1}}{(n-1)!} e^{-\mu x}, \quad x > 0$$

$$F_X(x) := P\{X \le x\} = 1 - \sum_{i=0}^{n-1} \frac{(\mu x)^i}{i!} e^{-\mu x}$$

# Moments

$$X \sim \mathrm{Erl}(n,\mu), \quad \mu > 0$$

- Mean value: $E[X] = E[X_1] + \ldots + E[X_n] = n/\mu$
- Variance: $D^2[X] = D^2[X_1] + \ldots + D^2[X_n] = n/\mu^2$
- Second moment: $E[X^2] = E[X]^2 + D^2[X] = n(n+1)/\mu^2$
- Standard deviation: $D[X] = \sqrt{D^2[X]} = (\sqrt{n})/\mu$
- Coefficient of variation: $C[X] = D[X]/E[X] = 1/(\sqrt{n}) \le 1$

# Mean residual lifetime

- Consider a random interval of length $X \sim \mathrm{Erl}(n,\mu)$.
  Assume that we know that the interval is longer than $x$.
  What is the mean residual lifetime?

$$\mathrm{MRL}(x) := E[X - x \mid X > x]$$

$$= \frac{\int_x^{\infty}(1 - F_X(y))\,dy}{1 - F_X(x)} = \frac{1}{\mu}\cdot\frac{\sum_{i=0}^{n-1}(n-i)\frac{(\mu x)^i}{i!}}{\sum_{i=0}^{n-1}\frac{(\mu x)^i}{i!}}$$

- The mean residual lifetime function $\mathrm{MRL}(x)$ is in this case decreasing (starting from $n/\mu$ and approching $1/\mu$)

# Hazard rate

- Consider a random interval of length $X \sim \mathrm{Erl}(n,\mu)$.
  Assume that we know that the interval is longer than $x$.
  What is the probability that it will end in a short interval of length $h$ after time $x$?

$$P\{X \le x+h \mid X > x\} = \frac{P\{x<X \le x+h\}}{P\{X>x\}} = \frac{f_X(x)h+o(h)}{1-F_X(x)}$$

- Thus, the hazard rate is

$$h(x) := \lim_{h\to 0} \frac{1}{h} P\{X \le x+h \mid X > x\} = \frac{f_X(x)}{1-F_X(x)} = \mu \cdot \frac{\frac{(\mu x)^{n-1}}{(n-1)!}}{\sum_{i=0}^{n-1} \frac{(\mu x)^i}{i!}}$$

- The hazard rate function $h(x)$ is in this case increasing (starting from $0$ and approching $\mu$)

# Hyperexponential distribution

$$X \sim \mathrm{Hyp}(n, p_1, \mu_1, \ldots, p_n, \mu_n), \quad \mu_i > 0, \; p_i > 0, \; \sum_i p_i = 1$$

- IID exponential phases in parallel; $X = I_1 X_1 + \ldots + I_n X_n$ where $X_i \sim \mathrm{Exp}(\mu_i)$ and $I_i \sim \mathrm{Bernoulli}(p_i)$ with $I_1 + \ldots + I_n = 1$
- $n$ = total number of phases
- $\mu_i$ = intensity of phase $i$, $p_i$ = probability of phase $i$

- Value space: $S_X = (0,\infty)$
- PDF and CDF:

$$f_X(x) = \sum_{i=1}^{n} p_i \mu_i e^{-\mu_i x}, \quad x > 0$$

$$F_X(x) := P\{X \le x\} = \sum_{i=1}^{n} p_i (1 - e^{-\mu_i x})$$

# Moments

$$X \sim \mathrm{Hyp}(n, p_1, \mu_1, \ldots, p_n, \mu_n), \quad \mu_i > 0, \;\; p_i > 0, \;\; \sum_i p_i = 1$$

- Mean value: $E[X] = E[I_1 X_1] + \ldots + E[I_n X_n] = p_1/\mu_1 + \ldots + p_n/\mu_n$
- 2nd moment: $E[X^2] = E[I_1 X_1{}^2] + \ldots + E[I_n X_n{}^2] = 2p_1/\mu_1{}^2 + \ldots + 2p_n/\mu_n{}^2$
- Variance: $D^2[X] = E[X^2] - E[X]^2 = \ldots$
- Standard deviation: $D[X] = \sqrt{D^2[X]} = \ldots$
- Coefficient of variation: $C[X] = D[X]/E[X] = \ldots \geq 1$

# Mean residual lifetime

- Consider a random interval with length $X \sim \mathrm{Hyp}(n, p_1, \mu_1, \ldots, p_n, \mu_n)$. Assume that we know that the interval is longer than $x$. The mean residual lifetime is now

$$\mathrm{MRL}(x) := \frac{\int_x^\infty (1 - F_X(y))\,dy}{1 - F_X(x)} = \frac{\sum_{i=1}^n p_i \frac{1}{\mu_i} e^{-\mu_i x}}{\sum_{i=1}^n p_i e^{-\mu_i x}}$$

- The mean residual lifetime function $\mathrm{MRL}(x)$ is in this case increasing (starting from $p_1/\mu_1 + \ldots + p_n/\mu_n$ and approching $\max_i 1/\mu_i$)

# Hazard rate

- Consider a random interval with length $X \sim \mathrm{Hyp}(n, p_1, \mu_1, \ldots, p_n, \mu_n)$. Assume that we know that the interval is longer than $x$. The hazard rate is now

$$h(x) := \frac{f_X(x)}{1 - F_X(x)} = \frac{\sum_{i=1}^{n} p_i \mu_i e^{-\mu_i x}}{\sum_{i=1}^{n} p_i e^{-\mu_i x}}$$

- The hazard rate function $h(x)$ is in this case decreasing (starting from $p_1\mu_1 + \ldots + p_n\mu_n$ and approching $\min_i \mu_i$)

# Pareto distribution

$$X \sim \text{Pareto}(b, \beta), \quad b > 0, \beta > 1$$

- heavy tail distribution
- $b$ = location parameter
- $\beta$ = shape parameter

- Value space: $S_X = (0, \infty)$
- PDF and CDF:

$$f_X(x) = \beta b \left(\frac{1}{1+bx}\right)^{\beta+1}, \quad x > 0$$

$$F_X(x) := P\{X \leq x\} = 1 - \left(\frac{1}{1+bx}\right)^{\beta}$$

# Moments

$$X \sim \text{Pareto}(b, \beta), \quad b > 0, \beta > 1$$

- Mean value: $E[X] = \int_0^\infty \beta b x (1 + bx)^{-\beta-1}\, dx = 1/(b(\beta - 1))$ for $\beta > 1$
- Second moment: $E[X^2] = \ldots = 2/(b^2(\beta - 1)(\beta - 2))$ for $\beta > 2$
- Variance: $D^2[X] = \beta/(b^2(\beta - 1)^2(\beta - 2))$ for $\beta > 2$
- Standard deviation: $D[X] = \sqrt{\beta}/(b(\beta - 1)\sqrt{(\beta - 2)})$ for $\beta > 2$
- Coefficient of variation: $C[X] = \ldots = \sqrt{\beta}/\sqrt{(\beta - 2)} \geq 1$ for $\beta > 2$

# Mean residual lifetime

- Consider a random interval with length $X \sim \text{Pareto}(b, \beta)$.
  Assume that we know that the interval is longer than $x$.
  The mean residual lifetime is now

$$\mathrm{MRL}(x) := \frac{\int_x^\infty (1 - F_X(y))\, dy}{1 - F_X(x)} = \frac{1 + bx}{b(\beta - 1)}$$

- The mean residual lifetime function $\mathrm{MRL}(x)$ is in this case linearly increasing (starting from $1/(b(\beta - 1))$ and approching $\infty$)

# Hazard rate

- Consider a random interval with length $X \sim \text{Pareto}(b, \beta)$.
  Assume that we know that the interval is longer than $x$.
  The hazard rate is now

$$h(x) := \frac{f_X(x)}{1 - F_X(x)} = \frac{b\beta}{1 + bx}$$

- The hazard rate function $h(x)$ is in this case decreasing (starting from $b\beta$ and approching $0$)

# Contents

- Basic concepts
- Discrete random variables
- Conditional expectation and variance
- Discrete distributions (count distributions)
- Continuous random variables
- Continuous distributions (time distributions)
- Other distributions and random variables

# Uniform distribution

$$X \sim \mathrm{U}(a,b), \quad a<b$$

- continuous counterpart of “casting a dice”

- Value space: $S_X = (a,b)$
- PDF:

$$f_X(x) = \frac{1}{b-a}, \quad x \in (a,b)$$

- CDF:

$$F_X(x) := P\{X \le x\} = \frac{x-a}{b-a}, \quad x \in (a,b)$$

- Mean value: $E[X] = \int_a^b x/(b-a)\,dx = (a+b)/2$
- Second moment: $E[X^2] = \int_a^b x^2/(b-a)\,dx = (a^2+ab+b^2)/3$
- Variance: $D^2[X] = E[X^2] - E[X]^2 = (b-a)^2/12$

# Standard normal (Gaussian) distribution

$$X \sim \mathrm{N}(0,1)$$

– limit of the "normalized" sum of IID r.v.s with mean $0$ and variance $1$

- Value space: $S_X = \Re$
- PDF:

$$f_X(x) = \varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- CDF:

$$F_X(x) := P\{X \le x\} = \Phi(x) := \int_{-\infty}^{x} \varphi(y)\, dy$$

- Mean value: $E[X] = 0$
- Variance: $D^2[X] = 1$

# Normal (Gaussian) distribution

$$X \sim \mathrm{N}(\mu, \sigma^2), \quad \mu \in \Re, \quad \sigma > 0$$

- if $(X-\mu)/\sigma \sim \mathrm{N}(0,1)$

- Value set: $S_X = \Re$
- PDF:

$$f_X(x) = F_X'(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

- CDF:

$$F_X(x) := P\{X \le x\} = P\left\{\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- Mean value: $E[X] = \mu + \sigma E[(X-\mu)/\sigma] = \mu$
- Variance: $D^2[X] = \sigma^2 D^2[(X-\mu)/\sigma] = \sigma^2$

# Properties

- ($i$) Linear transformation: Let $X \sim \mathrm{N}(\mu,\sigma^2)$ and $\alpha,\beta \in \Re$. Then

$$Y := \alpha X + \beta \sim \mathrm{N}(\alpha\mu + \beta, \alpha^2\sigma^2)$$

- ($ii$) Sum: Let $X_1 \sim \mathrm{N}(\mu_1,\sigma_1^2)$ and $X_2 \sim \mathrm{N}(\mu_2,\sigma_2^2)$ be independent. Then

$$X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

- ($iii$) Sample mean: Let $X_i \sim \mathrm{N}(\mu,\sigma^2)$, $i = 1,\ldots,n$, be IID. Then

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}(\mu, \tfrac{1}{n}\sigma^2)$$

# Central limit theorem (CLT)

- Let $X_1, \ldots, X_n$ be IID with mean $\mu$ and variance $\sigma^2$ (and the third moment exists)
- Central limit theorem (CLT):

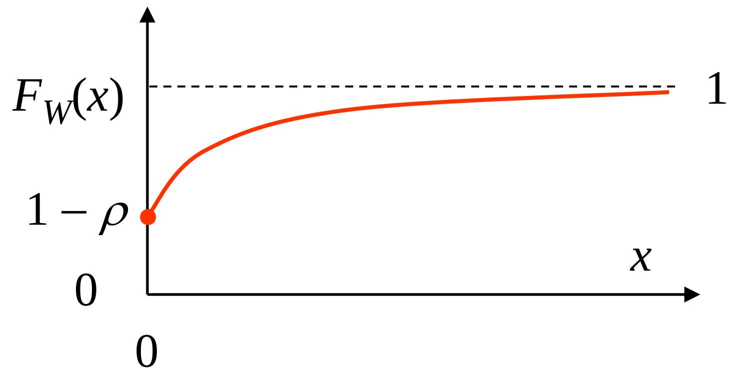
$$\frac{1}{\sigma / \sqrt{n}} (\bar{X}_n - \mu) \xrightarrow{\text{i.d.}} \mathrm{N}(0,1)$$

- It follows that for large values of $n$

$$\bar{X}_n \approx \mathrm{N}(\mu, \tfrac{1}{n}\sigma^2)$$

# Other random variables

- In addition to discrete and continuous random variables, there are so called mixed random variables
  - containing some discrete as well as continuous portions
- Example:
  - The customer waiting time $W$ in an M/M/1 queue has an atom at zero ($P\{W=0\}=1-\rho>0$) but otherwise the distribution is continuous



# Summary

- Basic concepts
  - Probability, conditional probability, independence, random variable, indicator, distribution, cumulative distribution function
- Discrete random variables
  - Point probabilities, expectation, variance, coefficient of variation
- Conditional expectation and variance
  - Conditioning rule, random sum of random variables, Wald's equation
- Discrete distributions (count distributions)
  - Bernoulli($p$), Bin($n,p$), Geom($p$), Poisson($a$)
- Continuous random variables
  - Density function, expectation
- Continuous distributions (time distributions)
  - Exp($\mu$), memoryless property, phase-type distributions, hazard rate, MRL

# Appendix: Useful formulas (1)

- Geometric sum:

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}, \quad 0 < x < 1$$

- Exponential function (1):

$$\sum_{i=0}^{\infty} \frac{x^i}{i!} = e^x, \quad x \in \Re$$

- Exponential function (2):

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x, \quad x \in \Re$$

# Appendix: Useful formulas (2)

- Binomial theorem:

$$(a+b)^m = \sum_{n=0}^{m} \frac{m!}{n!\,(m-n)!} a^n b^{m-n}$$

- Multinomial theorem:

$$(a_1 + \cdots + a_k)^m = \sum_{n \in S_m} \frac{m!}{n_1! \cdots n_k!} a_1^{n_1} \cdots a_k^{n_k}$$

$$S_m := \{n = (n_1, \cdots, n_k) \geq 0 \mid n_1 + \cdots + n_k = m\}$$